AD-A258 553

WL-TR-92-3095

MODELING THE DYNAMICS OF MENTAL
WORKLOAD AND HUMAN PERFORMANCE IN
COMPLEX SYSTEMS

WILLIAM B. ROUSE
SHARON L. EDWARDS
JOHN M. HAMMER
SEARCH TECHNOLOGY
4725 PEACHTREE CORNERS CIRCLE
NORCROSS, GEORGIA   30092

AUG 1992

FINAL REPORT FOR 09/01/91 - 08/10/92

DTIC
ELECTE
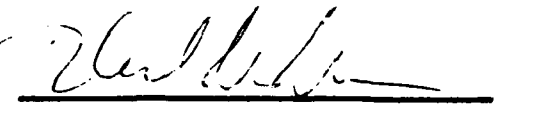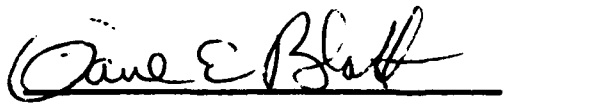DEC 15 1992
S E D

92-31388

92 12 14 068

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

HARRY W. VANDERBACH, 1Lt, USAF
Contract Monitor
Cockpit Development Branch

RICHARD W. MOSS, GM-14
Chief, Cockpit Development Branch
Cockpit Integration Division

PAUL E. BLATT
Chief, Cockpit Integration Division
Flight Dynamics Directorate

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify WL/FIPC , WPAFB, OH 45433-6553 to help us maintain a current mailing list.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE AUG 1992	3. REPORT TYPE AND DATES COVERED FINAL 09/01/91--08/10/92

4. TITLE AND SUBTITLE MODELING THE DYNAMICS OF MENTAL WORKLOAD AND HUMAN PERFORMANCE IN COMPLEX SYSTEMS

6. AUTHOR WILLIAM B. ROUSE
SHARON L. EDWARDS
JOHN M. HAMMER

5. FUNDING NUMBERS

C F33615-88-C-3612
PE 62201
PR 2403
TA 04
WU 98

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SEARCH TECHNOLOGY
4725 PEACHTREE CORNERS CIRCLE
NORCROSS, GEORGIA 30092

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

WRIGHT LABORATORY
AIR FORCE SYSTEMS COMMAND
WRIGHT PATTERSON AFB OH 45433-6553
WL/FIPC, Attn: VANDERBACH 513-2556670

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

WL-TR-92-3095

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This program studied the relationship between subjective workload and human behavior and proposed a model of the dynamics of this relationship. Results of three simulation experiments are detailed in this report and show that simple linear indentification algorithms are robust in online identification of noisy, nonlinear versions of the model. This model and the associated algorithms have the potential to enable online inferences of workload and could be used to prompt/invoke human aiding or automated systems to help reduce workload. Applications for such systems exist in aiding aircraft pilots; command, control, communication decision makers; and other personnel in dynamic, time constrained environments.

14. SUBJECT TERMS WORKLOAD, HUMAN PERFORMANCE, PILOT AIDING, SUBJECTIVE WORKLOAD, CONTROL MODELS, WORKLOAD MODELS.

15. NUMBER OF PAGES 30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

SUMMARY

The relationship between subjective workload and subsequent human behaviors is considered. A model of the dynamics of this relationship is proposed. Results of three simulation experiments are reported that show that simple linear identification algorithms are surprisingly robust in online identification of noisy, nonlinear versions of this model. The implications for developing adaptive aiding and decision support systems are noted.

The results reported here have important potential implications for the Air Force. The model discussed offers a possible new means of dealing with the problem of workload in a very practical way. The software package developed for evaluating this model is likely to be of use to other researchers for assessing the breadth of applicability of this model. The simulation results available to date support the potential broad applicability of this model to Air Force problems. In particular, the model and associated algorithms could be embedded in software for aiding aircraft pilots, C3 decision makers, and other personnel in dynamic, time-constrained environments. Such applications would enable online inferences of workload and, if excessive, prompt and/or invoke aiding or automation to reduce workload to acceptable levels.

DTIC QU~ ~.

Accesion For		
NTIS CRA&I		☒
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

INTRODUCTION

The concept of mental workload has long been recognized as an important, albeit elusive, factor in human performance in complex systems (Moray, 1979; Gopher & Donchin, 1986; O'Donnell & Eggemeir, 1986). Much research has been devoted to defining workload constructs and conducting careful measurements. However, the "holy grail" has not been found, and it has been argued that approaches to workload should be reconsidered (Hart, 1988).

What is the difficulty? This question can be considered by discussing a set of finer-grained questions:

- What is workload?
- Why does it matter?
- How can workload be measured?
- How can it be predicted?
- How can it be controlled?

The first question can be answered in several ways. One answer is that workload is the demand placed upon humans. The usefulness of this definition depends on the extent to which humans attempt to satisfy all demands, a condition which is often satisfied in experiments but seldom satisfied in real life.

While demand relates to the load imposed on people, the load experienced may be a better indicator of workload in the sense that it reflects the human information processing that actually occurs, rather than what was demanded. Use of information processing resources may be indicative of capacity utilized and, by inference, capacity remaining.

This perspective leads immediately to the problem of measuring the use of information processing resources. Just as constructs such as energy lead to "proxy" measures such as temperature and pressure, workload in terms of resource utilization must be measured using proxy measures. Thus, workload researchers measure, for example, skin resistance, heart rhythms, and brain potentials. Changes of such measures are said to reflect changes of workload if two tasks with arguably different levels of workload yield different values of proxy measures.

A third approach to workload focuses on people's perceptions of their own workload, typically measured by their reports of their subjective experience of workload. It can be argued, and often is, that subjective reports do not necessarily reflect objective workload. For example, people may rate a task as not requiring much effort, but in fact utilize near maximum information processing resources to do the task.

People may be sufficiently skilled to perform a difficult task almost "effortlessly." Similarly, a seemingly very simple task may completely stymie many people due to lack of knowledge and skills. Thus, workload is necessarily task specific and person specific.

Choosing among the above three approaches to workload depends on why workload matters. One reason to be concerned with workload is the possibility that excessive workload will result in people being unable to do their tasks, or at least not able to do them well enough. A related concern is the possibility that people may be able to do their task acceptably but, due to excessive workload, may not be able to persist or perhaps even survive over a long period of time.

These two concerns typically prompt a desire to measure the objective workload -- in terms of either load imposed or load experienced -- of particular tasks

and, if consequences such as outlined above are likely, tasks are redesigned. Alternatively, aiding and/or training might be provided. The result is that tasks are "certified" as doable, at least in terms of workload.

Another type of concern -- the concern in this report -- is people's reactions to workload. In other words, how do people behave in high workload situations? Several recent efforts (Hart, 1988; Morris & Rouse, 1988) have indicated that people do not simply put up with high workload. They adapt their behaviors to cope with the workload by, for example, changing strategies, delaying low priority tasks, and accepting degraded but satisfactory performance.

It seems reasonable to argue that these types of adaptive behaviors are in response to perceptions of workload. If these perceptions are verbalized, although they usually are not, then they might be recorded as subjective workload ratings. Such ratings are proxy measures of people's perceptions. The use of such measures to explain and predict behaviors is the focus of this report.

Before pursuing this topic in depth, it is important to discuss the distinction between the effects of workload on behavior and its effects on performance. Traditional workload research has attempted to relate some statistical measure of workload (e.g., the average) to some statistical measure of performance (e.g., again, the average). The basic hypothesis is that high workload, and perhaps low workload, will result in degradations of how well tasks are done. Despite numerous studies of this hypothesis, it is usually concluded that there is insufficient evidence to reject the null hypothesis that workload and performance are unrelated.

Why does such a counterintuitive result occur so frequently? The answer, it can be argued, is that people can respond to workload excesses, or deficits, by changing what they do, but not necessarily how well. In other words, they can

3

change their behaviors, which may or may not affect performance. Thus, workload will not affect performance unless changes of behavior affect performance, which may be difficult to assess in realistically complex tasks -- see, for example, Enstrom & Rouse (1977) where root-mean-squared tracking error was the least sensitive measure (of nine) to distractions in manual control tasks.

A further complication with traditional studies is the typical reliance on aggregate measures. If people change their behavior in response to workload changes, any performance impact of these behavioral changes occurs later in time since behavior seldom affects performance measures instantaneously. To capture this phenomenon, analyses must relate moment-to-moment variations of workload to moment-to-moment changes of behavior which, in turn, relate to subsequent moment-to-moment variations of performance. This requires the use of time series analysis, or equivalent, rather than traditional analysis of variance methods.

Two recent studies of workload in process control tasks employed time series analysis to estimate autocorrelations and cross-correlations of subjective workload, behavior, and performance (Morris & Rouse, 1988). Substantial time-lagged relationships were found among moment-to-moment variations of workload, behavior, and performance. The obvious conclusion is that subjectively reported mental workload is a dynamic process, rather than a static property of task situations. Thus, the average workload, as well as the average performance, are inadequate measures -- at the very least, averages provide poor means for understanding mental workload.

This report focuses on the dynamics of mental workload. A model is formulated that depicts workload dynamics in the context of system control. The identifiability of relationships within this model is discussed, with emphasis on the

possibility of online identification for the purpose of estimating and predicting workload and behavior. Results of a simulation experiment are reported that illustrate the sensitivity of identification accuracy to a variety of parameters. These results show that very simple linear identification methods are somewhat surprisingly robust in noisy, nonlinear environments. It is concluded that these results provide initial evidence that the adaptability of performance aids and decision support systems can be substantially enhanced.

The results reported here have important potential implications for the Air Force. The model discussed in the next section offers a possible new means of dealing with the problem of workload in a very practical way. The software package developed for evaluating this model is likely to be of use to other researchers for assessing the breadth of applicability of this model. The simulation results available to date, and presented later in this report, support the potential broad applicability of this model to Air Force problems. In particular, the model and associated algorithms could be embedded in software for aiding aircraft pilots, C3 decision makers, and other personnel in dynamic, time-constrained environments. Such applications would enable online inferences of workload and, if excessive, prompt and/or invoke aiding or automation to reduce workload to acceptable levels.

THE MODEL

The variables within the model include four vectors:

$x(t)$ = system state at time t
$u(t)$ = human activity at time t
$p(t)$ = system performance at time t
$s(t)$ = subjective experience of workload at time t

As indicated in Figure 1, subjective experience (s) is related to what people do (u), how the system responds (x), and perhaps how well performance objectives are

achieved (p)[1]. However, performance is only related to subjective experience indirectly, via two transformations -- even feedback of performance results (shown dotted in Figure 1) is typically delayed and, therefore, cannot affect behavior directly. A more direct research question is how subjective experience s affects behavior u -- as emphasized earlier, this is the concern in this report.
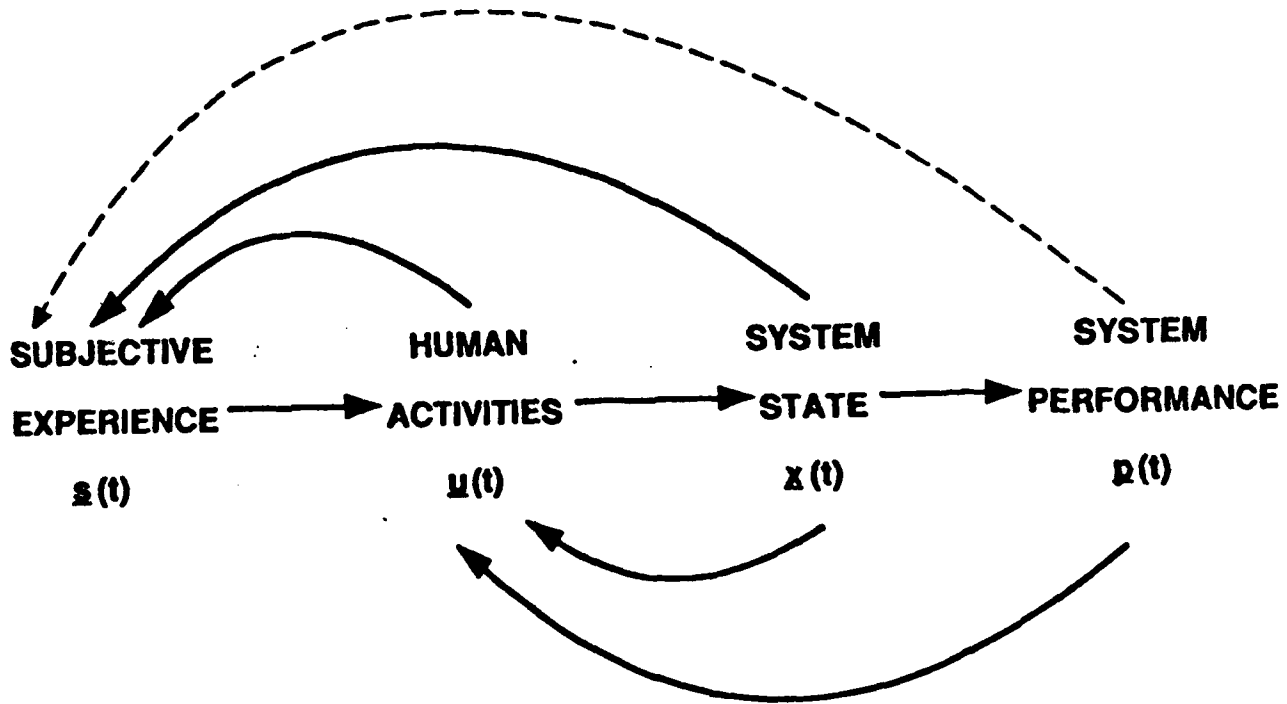

Figure 1. Basic Relationships

Based on the relationships shown in Figure 1, the elements of the model can be explicated. The next state of the system depends on the current state $x(t)$ and current inputs $u(t)$, i.e.,

$$x(t+1) = F[x(t), u(t)].\tag{1}$$

---

[1] These variables might be expressed in different terms, e.g., (Gopher & Braune, 1984), but such characterizations tend not to capture the time variations of interest.

Note that relationship (1) specifies a dynamic system but leaves open the form of the function F -- it could be of standard mathematical form or it could specify logical relationships within a rule-based system. Quite simply, F describes how the system (e.g., an aircraft) dynamically responds and typically is known or available from the designers of the system.

The "subjective state" of the human depends on the state of the system, human's inputs and previous subjective experience, i.e.,

$$\underline{s}(t+1) = G[\underline{x}(t), \underline{u}(t), \underline{s}(t)].$$ (2)

Thus, the evolution of subjective experience $\underline{s}(t)$ is represented as a dynamic process which allows for carryover effects, e.g., "psychological inertia" due to recent high workload, or stress due to "operating on the edge" as found in Morris & Rouse (1988). Equation (2) also allows the possibility of changes of behavior $\underline{u}(t)$ to attenuate subjective experience $\underline{s}(t)$ -- in other words, adaptive strategies and behaviors.

Defining the augmented state of the human-machine system as the vector containing both the system state and the human's perception of his/her own state,

$$\underline{x}^*(t+1) = \begin{bmatrix} \underline{x}(t+1) \\ -------- \\ \underline{s}(t+1) \end{bmatrix}$$ (3)

enables combining relationships (1) and (2) to yield

$$\underline{x}^*(t+1) = F^*[\underline{x}^*(t), \underline{u}(t)]$$ (4)

where F^* is related to F and G in a mathematical and/or logical manner. Thus, the model assumes that the human's behavior serves the purpose of controlling the system state and his/her own state.

The relationship between the human's activity $\underline{u}$ and the augmented state can be characterized by

$$\underline{u}(t+1) = H[\underline{x}^*(t), \underline{u}(t)] \qquad (5)$$

which depicts $\underline{u}$ as evolving dynamically. This allows, for example, the possibility that $\Delta\underline{u}$ is limited -- that is, its rate of change is limited due to neuromotor dynamics.

Given $\underline{x}(t)$, $\underline{u}(t)$, $\underline{s}(t)$ for t = 1,2, ..., one can identify F* and H using one or more of a variety of identification algorithms (Rouse, et al., 1989). If relationship (1) represents a signal processing task (e.g., flight control), identification of F* and H should be reasonably straightforward. In contrast, if relationship (1) represents a symbol processing task (e.g., flight planning or re-planning), identification is possible although cumbersome. The algorithms used in this study are discussed in a later section.

A comparison of F (which is typically known or can be similarly identified) with F* enables determination of G, although this may not be a straightforward mathematical or logical process. With G and H known, it is possible to predict subjective experience $\underline{s}(t)$ via G, and also predict behavior $\underline{u}(t)$ via H.

USE OF THE MODEL

The capability to predict subjective experience $\underline{s}(t)$ and behavior $\underline{u}(t)$, via knowledge of G and H, can greatly enhance intelligent interfaces associated with aiding and decision support (Rouse, 1988; Rouse, et al., 1988, 1990; Rouse & Morris, 1987). Based on the above model, a support system can predict people's subjective experience and the consequent activities. Alternatively, subjective experience might be inferred from activities. Either possibility can be of great value

for support functions such as information management, adaptive aiding, and error monitoring.

For such a scheme to work, measurement of $\underline{x}(t)$, $\underline{u}(t)$, and $\underline{s}(t)$ must be possible. Measurement of system state $\underline{x}(t)$ and behavior $\underline{u}(t)$ is straightforward. Subjective experience $s(t)$ can be assessed in a variety of ways. Reasonable success has been achieved using auditory probes and oral responses on ten-point rating scales (Johannsen & Rouse, 1983; Morris & Rouse, 1988). It is easy to imagine alternative approaches that involve, for example, visual probes and manual responses such as pressure on a control element.

The most important factor in choosing among approaches for measuring subjective experience $\underline{s}(t)$ is the purpose that these measurements are serving. It is completely unnecessary to make the assertion that $\underline{s}(t)$ is synonymous with any of the traditional metrics of workload. It is also unnecessary for $\underline{s}(t)$ to be strongly related to performance $\underline{p}(t)$. All that is necessary is that subjective experience $\underline{s}(t)$ affect strategies and behaviors that directly influence behavior $\underline{u}(t)$. Fortunately, as noted earlier, there is strong evidence that $\underline{s}(t)$ and $\underline{u}(t)$ are related (Hart, 1988; Morris & Rouse, 1988).

The frequency with which subjective experience $\underline{s}(t)$ can be measured does pose a potential problem since $\underline{s}(t)$ can only be sampled, say with period T. Given measurements of $\underline{x}(t)$ and $\underline{u}(t)$ for $t = 1, 2, ...,$ and $\underline{s}(t)$ for $t = T, 2T, ...,$ the essence of the identification problem is determining the best-fit relationship F* that enables prediction of subjective experience $\underline{s}(t)$ and, hence, behavior $\underline{u}(t)$ for $t = 1, 2, ...$ If T = 1, the sampling problem is obviously eliminated. However, people cannot be probed at this rate. More realistically, T could be 10-20 for experimental purposes, and perhaps 50-100 for calibration purposes within applications.

9

This constraint can be handled in two ways. One approach is to interpolate between samples of $\underline{s}(t)$. The simplest form of this approach would be linear interpolation. In other words, a straight line could be drawn between $\underline{s}(T)$ and $\underline{s}(2T)$, for example, and this line used to estimate $\underline{s}(T+1)$, $\underline{s}(T+2)$, ... This is one of the most rudimentary forms of interpolation.

A more sophisticated approach is to model subjective experience $\underline{s}(t)$ as a discrete time series with time units equal to T. The resulting model can then be converted to continuous time, and then transformed back to discrete time for T=1. While this does not recreate information lost due to sampling, it does provide a much more rigorous approach to dealing with missing data than possible with simple interpolation.

This problem can be minimized, but not eliminated, by using as small a value of T as possible during model development. If the resulting F^* is reasonably robust, across conditions and people, it can form the basis for model-based estimation (interpolation) in situations where T cannot be small. If, on the other hand, F^* is totally situation and person dependent, then the proposed approach is unlikely to be very successful. The essence of the problem, therefore, is whether or not F^* can be identified (feasibility) and whether or not these methods can be applied in realistic settings (practicability).

A final issue concerns how the ability to predict subjective experience $\underline{s}(t)$ and behavior $\underline{u}(t)$ would be integrated into an aiding scheme, as well as how the nature of aiding would affect $\underline{s}(t)$ and, consequently, $\underline{u}(t)$. The basic idea is to use $\underline{s}(t)$ probes until G and H are identified, and then only occasionally use $\underline{s}(t)$ probes for calibration. The stability and universality of G will strongly influence this process. The structure of G will have to be sufficiently robust to handle the types of events

noted earlier. The parameters of G will have to be sufficiently stable to enable periods of open-loop predictions of at least 50-100 seconds in length, and hopefully many minutes. G may have to be tailored to each person -- hopefully via parametric rather than structural changes.

SIMULATION EXPERIMENTS

To explore the feasibility of the proposed model of the dynamics of workload, three simulation experiments were performed. The simulation environment used is depicted in Figure 2.
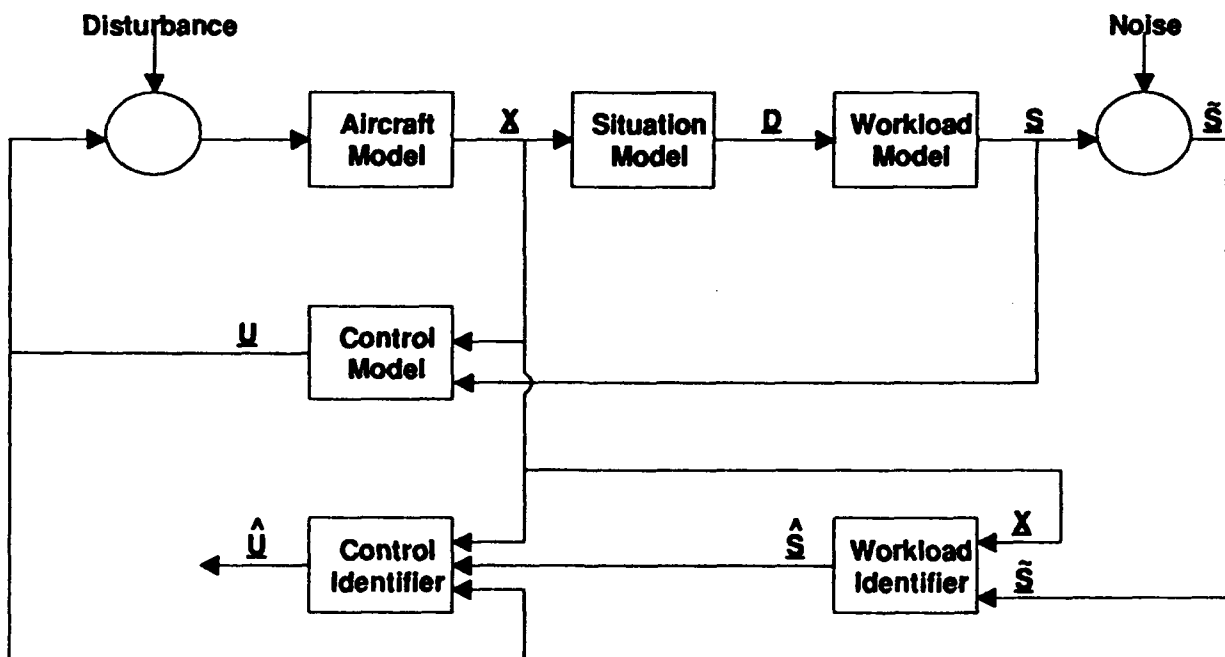
Figure 2. Simulation Environment

11

The task simulated was compensatory control of the role and pitch of an aircraft modeled as a second order linear dynamic system, characterized by natural frequencies and damping ratios indicated in later discussion. The situation model mapped horizontal and vertical deviations to a single distance metric.

Two different workload models were employed and are described in the discussion of each experiment. Two control models were also used, one a standard optimal linear controller and the other a linear controller whose gains were affected by workload as discussed below. Linear models of manual control are among the few human performance models that can be used as "off-the-shelf" components of larger models -- see Chapter 3 of Rouse (1980) for a review of the background and nature of these models.

Note that these experiments did <u>not</u> involve human controllers. Humans' perceptions of workload and control behavior were modeled. In this way, the experimenters always knew the "ground truth" which was then compared to the outputs of the identifiers.

It is also important to note that the simple task chosen provided a strong test of the proposed model in the following way. Since the baseline for all of the models indicated in Figure 2 is linear or near linear, linear estimation theory provides a strong basis for identification. As nonlinearities and noise are added, as discussed below, linear identification methods will start to have problems. If degradation is abrupt, then we must conclude that the proposed model is much too fragile to be of practical value, particularly for more realistic aircraft, situation, and workload models.

Identification

Figure 2 indicates two identifiers, one for workload and the other for control, or G and H from equations 2 and 5, respectively. The simulation experiments described later in this section were designed to test these identifiers.

The purpose of the workload identifier is to determine relationships between observed states $\underline{x}$ and perceptions of workload corrupted by noise $\underline{\tilde{s}}$. The "ground truth" is that workload is related to the perceived situation. However, the identifier is not privy to this knowledge.

The workload identifier is also faced with the problem of only being able to sample $\underline{\tilde{s}}$ every T time units. As discussed in an earlier section, this problem can be handled by interpolation of some form or discrete-continuous-discrete transformations that enable estimating intervening, non-sampled values of $\underline{s}$.

The control identifier shown in Figure 2 is much more straightforward than the workload identifier. Its purpose is to determine how workload $\underline{\hat{s}}$, as well as the system state $\underline{x}$ and past control actions $\underline{u}$, affect future control actions. This is a simple linear identification problem, except for those cases where the "ground truth" control model involves a nonlinear modification of control gains due to workload.

For both identification problems, a simple linear identifier was used as depicted in Figure 3. A rudimentary view of how such an identifier works is

$$\text{New Model} = \text{Old Model} + \text{Gain (Observation - Prediction)} \qquad (6)$$

where the predicted output of the process of interest (i.e., workload ratings or control behaviors) is based on the old model which, as shown in (6), is then compared to the actual output of the process. A variation of this type of identifier is a fading-memory identifier where the gain is adjusted to enable, in effect, forgetting of older

13

observations. Derivation and two other types of application of linear identifiers are discussed in Enstrom and Rouse (1977) and Rouse (1977).
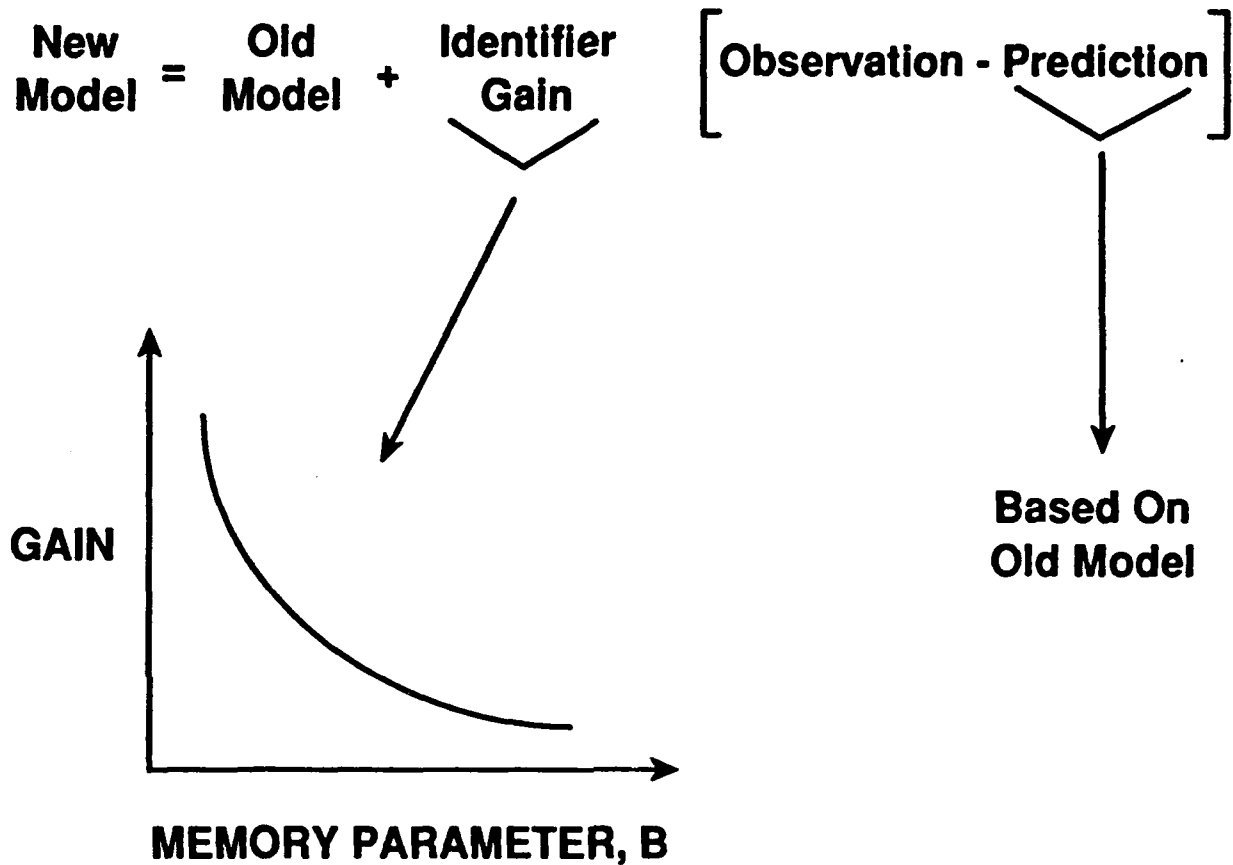
$$\text{New Model} = \text{Old Model} + \text{Identifier Gain} \left[\text{Observation - Prediction}\right]$$

GAIN

Based On Old Model

MEMORY PARAMETER, B

Figure 3. Approach to Identification

Experiment 1

Three second-order aircraft models were considered: underdamped, critically damped, and overdamped. This resulted in fast, but oscillatory responses to control inputs (underdamped); responses as fast as possible without oscillation (critically damped); and sluggish responses (overdamped). The natural frequency and damping ratio for each of these models are indicated in Figure 4a. The control model was the optimal linear control for each aircraft model.

Type of Model	Natural Frequency	Damping Ratio
Overdamped	0.5	8.0
Critically Damped	0.5	1.0
Underdamped	0.5	0.125

Figure 4a. Aircraft Model Parameters

Type of Model	Natural Frequency	Damping Ratio
Overdamped	0.05	4.0
Underdamped	0.05	0.25

Figure 4b. Linear Workload Model Parameters

Type of Model	Time Constant
Slow	10
Fast	1

Figure 4c. Heuristic Workload Model Parameters

A purely linear workload model was used for this experiment. A second order model was used based on the cross-correlation results (see Figure 5) for the time series analysis in Morris & Rouse (1988). Both underdamped (oscillatory response to situation changes) and overdamped (sluggish response) models were considered. The natural frequency and damping ratio for each of these models are shown in Figure 4b.
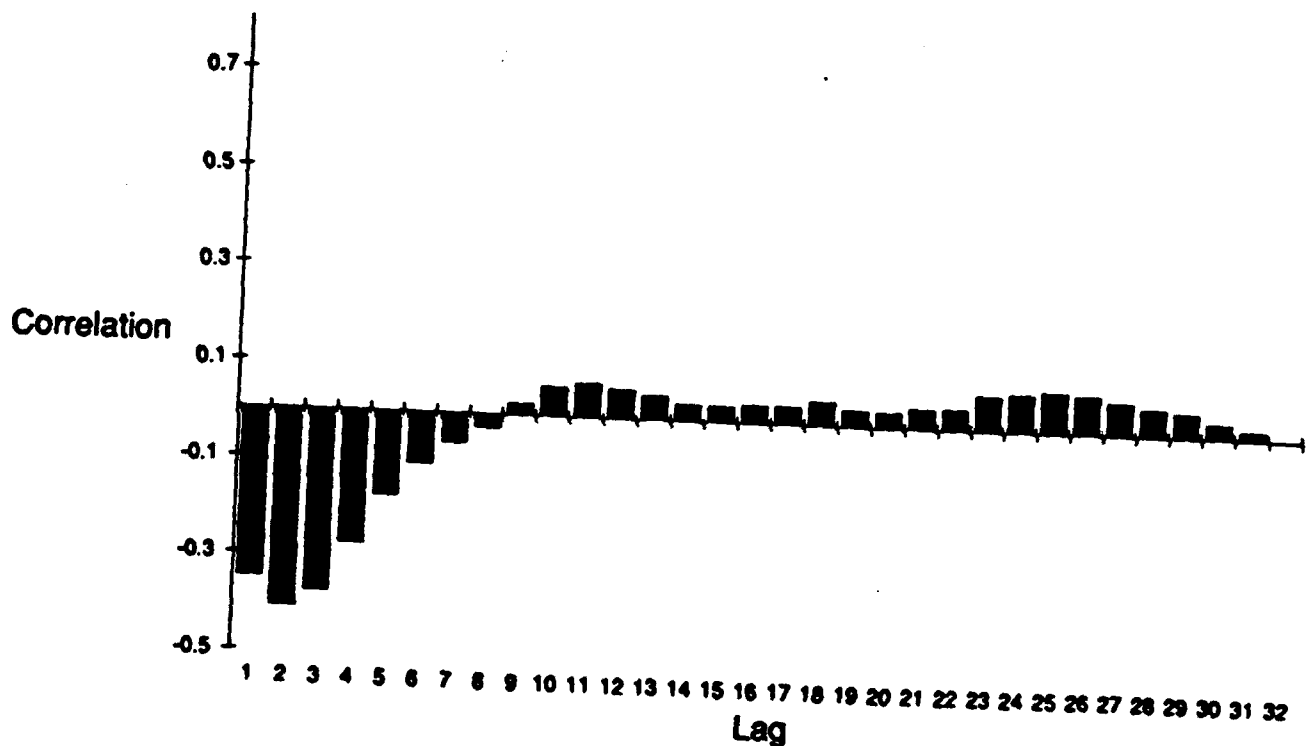
Figure 5. Cross-Correlation for Effort Ratings and Performance
(Morris & Rouse, 1988)

The primary independent variables in this experiment were the noise level on workload samples and the workload sampling period. Noise levels studied were 0, 10, and 20% where these percentages refer to the standard deviation of the noise as a percentage of the magnitude of $\underline{s}$. Research has indicated (Rouse, 1980) that error in judging a variable is typically roughly 10% of the magnitude of the variable. Sensitivity to this percentage was assessed by also considering 0% error and 20%,

or twice the typical level. Sampling periods studied included 1, 2, 5, 10, and 20 time units.

The dependent measure of interest was the root-mean-squared error in predicting workload, divided by the root-mean-squared magnitude of the workload. If this ratio is small, the proposed model is performing well. As this ratio approaches 1.0, one would get equally good results by simply predicting the mean workload rather than using the proposed model of the dynamics of workload.

Experiment 2

The second experiment considered the same parameter variations, but changed the workload model. The second-order linear model was replaced with a heuristic model based on results from Morris & Rouse (1988) -- see Figure 6.

This heuristic model was a first-order linear model driven heuristically by variations of the perceived situation, characterized by the nonlinear distance metric. These heuristics were based on the nature of the situation (acceptable vs. unacceptable) and its likely evolution (improving vs. worsening). Acceptability was determined by thresholds on the distance metric and its rate of change. Judgments regarding the likely evolution of the situation were based on rates of change of these variables. A negative rate of change indicated an improving situation (i.e., distance decreasing) while a positive rate of change indicated a worsening situation (i.e., distance increasing). Figure 7 summarizes the heuristic portion of this model indicating the value of workload ($\underline{s}$) for various positions in the $\underline{D}$ vs. $\underline{\dot{D}}$ plane.

Within this heuristic model, two variations of the first-order model parameter (i.e., the time constant) were studied to compare fast responding to slowly responding dynamics. The parameters used for this model are indicated in Figure
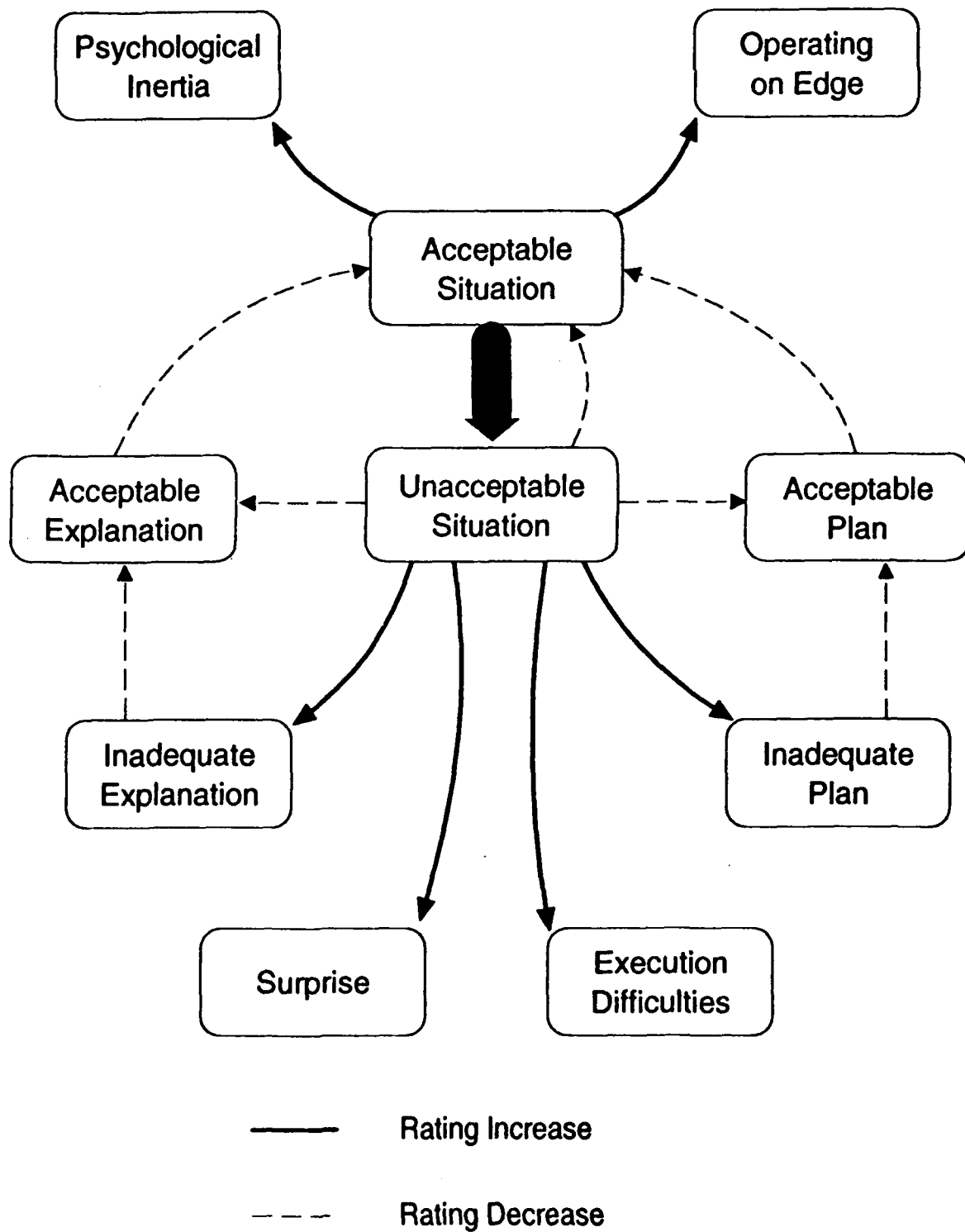
Figure 6. Factors Affecting Changes in Effort Ratings (Morris & Rouse, 1988)
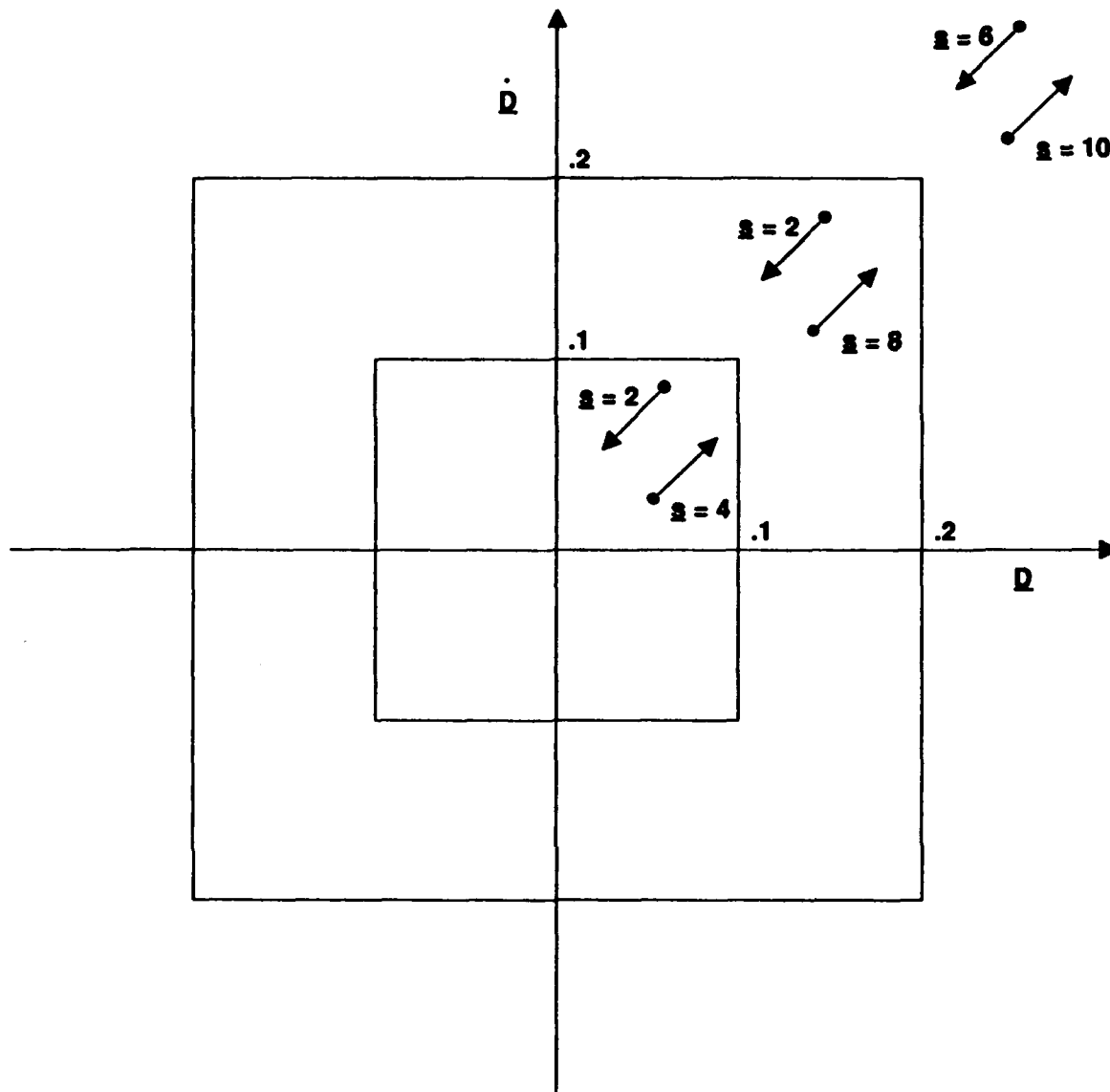
18

Figure 7. Heuristic Portion of Workload Model

4c. The combination of the heuristic portion of the model and the first-order component resulted in a nonlinear model, due to the discontinuous gain created by the heuristics, where workload changes responded either quickly (small time constant) or slowly (large time constant) to situational changes.

Experiment 3

This experiment focused on identifying the control model. All parameter variations were the same as those in Experiment 1. The one exception was the use of the heuristic control law whereby controller gains were modified by a multiplicative workload term. This term decreased gains for low workload, proportionally increased gains with increasing workload until moderate workload where gains were constant and unattenuated. For high workload, gains were proportionally decreased as workload increased, at a rate much smaller than that for low workload. This heuristic is based on the results reported in Morris and Rouse (1988).

The dependent measure for this experiment was the root-mean-squared error in predicting control actions, divided by the root-mean-squared magnitude of control. As with the workload-related ratio, if this ratio is small, the proposed model is performing well. As this ratio approaches 1.0, equally good results would be obtained by simply predicting the mean control.

RESULTS

All of the simulation results discussed in this section are based on simulation runs of 200 time units. All data points on the figures shown below are averages across 10 runs.

Two results hold across all experiments. First, rudimentary linear interpolation between samples does not work. The ratios of root-mean-squares became completely unacceptable -- much greater than 1.0. Consequently, all of the results discussed in this section are based on the aforementioned discrete-continuous-discrete transformations.

The second result that holds across all experiments is a lack of benefit from employing identification algorithms that assume "ground truth" to be linear with time-varying parameters. We had expected that a "fading memory" identifier would be better able to deal with the nonlinearities in the workload and/or control models. Results indicated, however, that the best rate of "fading" was no fading. In other words, there was no value in "forgetting" older observations and standard linear identification algorithms worked best.

Experiment 1

The results for the first experiment are shown in Figures 8a-8f. Not surprisingly, the linear identifier did quite well for linear workload model. As expected, the error was near-zero when the noise level was 0%. Small errors, nevertheless, arose due to the identifier being unaware of the transformation of x to the nonlinear distance metric.

Identification results were best for overdamped aircraft models and underdamped workload models. In other words, slowly varying x and rapidly responding s led to the best identification results.

Identification errors increased with sampling period T if there was noise. The rate of increase was related to the noise level. It appears in these plots that errors are roughly proportional to the product of sampling period and noise level.

For long sampling periods (not shown in figures), identification error starts to increase sharply. The "knee" in this curve is related to the nature of the dynamics. Beyond a certain sampling period, the identifier does not get enough real data to do a good job. This result agrees with sampling theory and, hence, was expected.
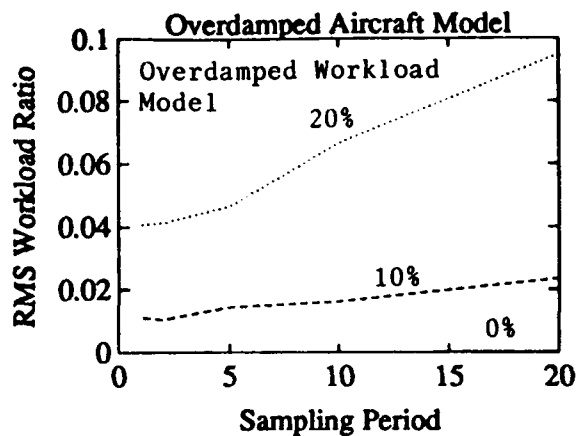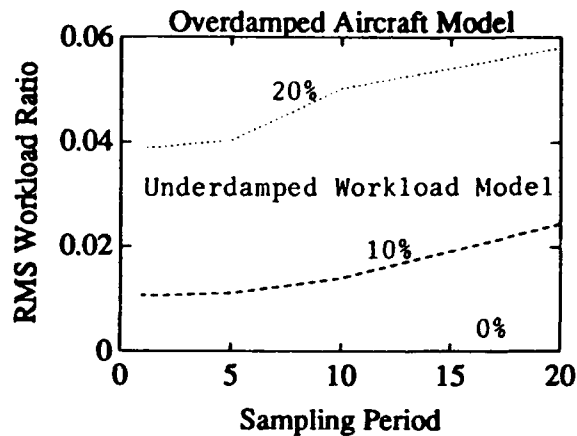
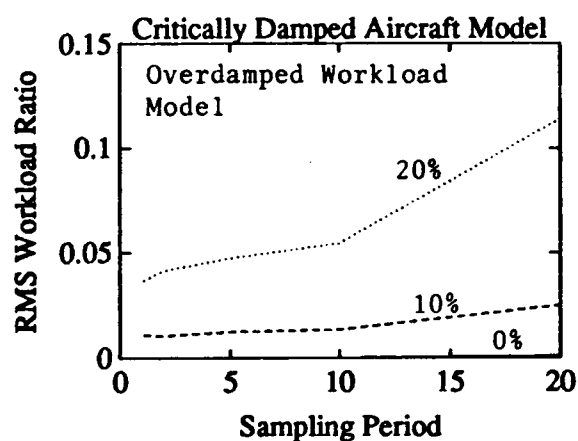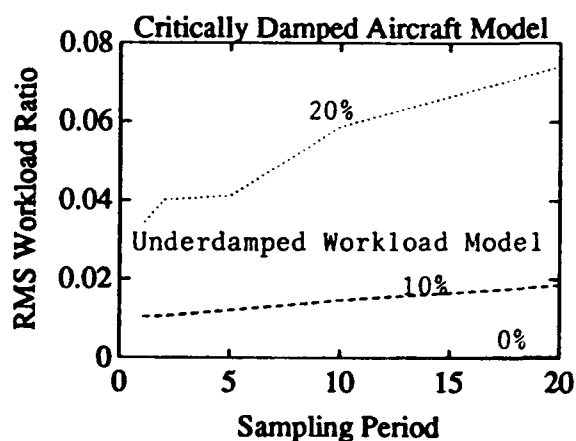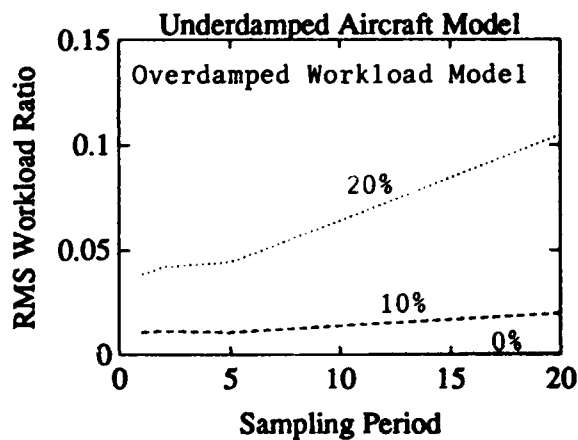Figure 8a

Figure 8d

Figure 8b
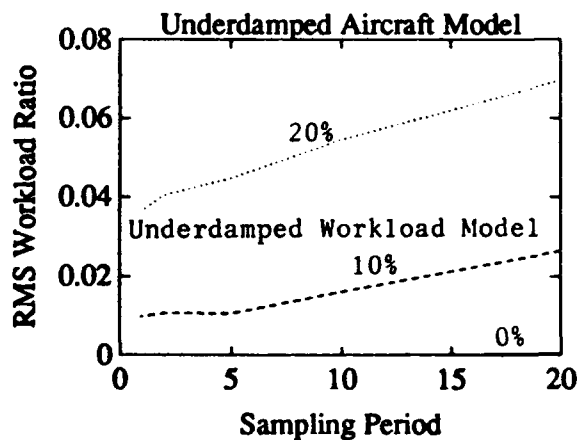

Figure 8e

Figure 8c

Figure 8f

Figure 8. Results for Identification of Linear Workload Model

22

Experiment 2

The results for the study of the heuristic model are shown in Figures 9a-9f. The fast response heuristic model (Figures 9a-9c) leads to the best results, because its response is more linear than the slow response model. The results for the fast heuristic model are comparable to those for the linear overdamped workload model. For longer sampling periods, performance degrades for the fast heuristic model in a manner comparable to that for the linear models.

Results for the slow heuristic model (Figures 9d-9f) were not quite as good as the fast model, which is not surprising because it is much less linear. However, the slow heuristic model is not as sensitive to sampling period. Thus, identification is not as good for the slow heuristic model but it also is not as fragile.

Experiment 3

The results for identification of the heuristic control model were all in the range of 25-30%, independent of noise level and independent of sampling period. This is not surprising because the "ground truth" control model did not, in itself, include linear dynamics. Thus, while the linear identifier could have found dynamics in the control model, if such existed, the model was not dynamic and sampling period was not a problem.

As noted earlier, the control model included a linear function of aircraft state $\underline{x}$, heuristically attenuated by workload $\underline{s}$. The linear identifier was able to find a reasonable linear relationship between $\underline{u}$ and the inputs $\underline{x}$ and $\hat{\underline{s}}$ despite the inherent nonlinearity of the underlying "ground truth."
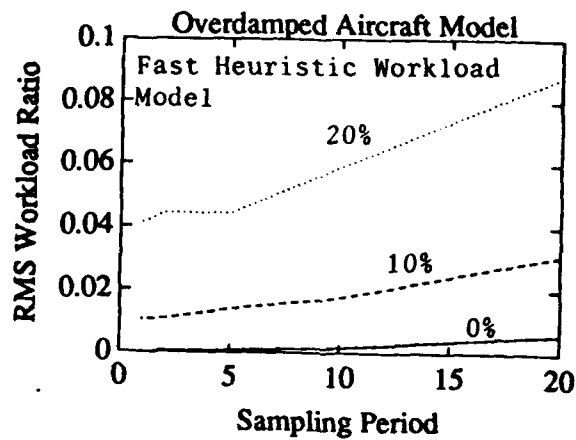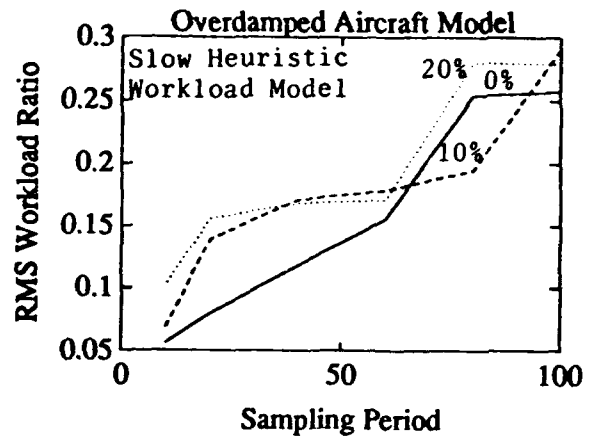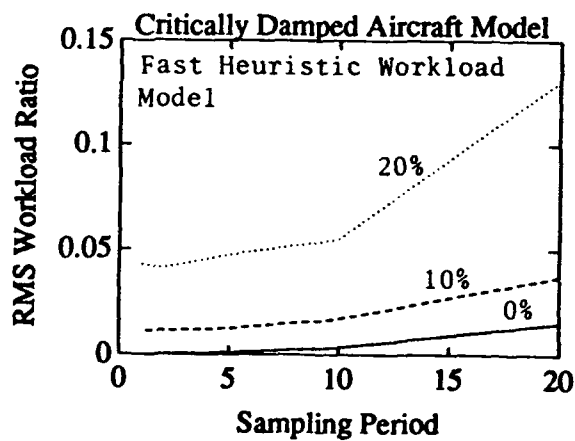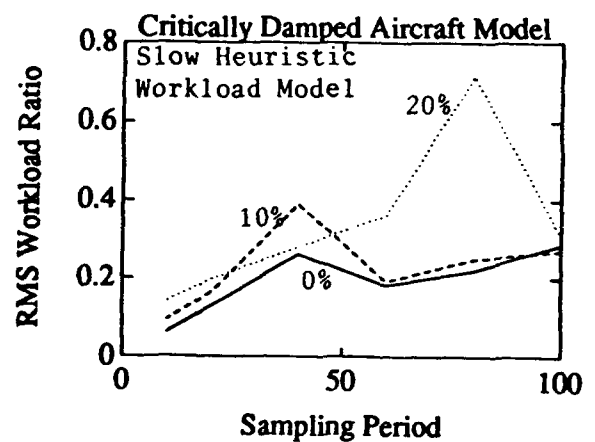
Figure 9a

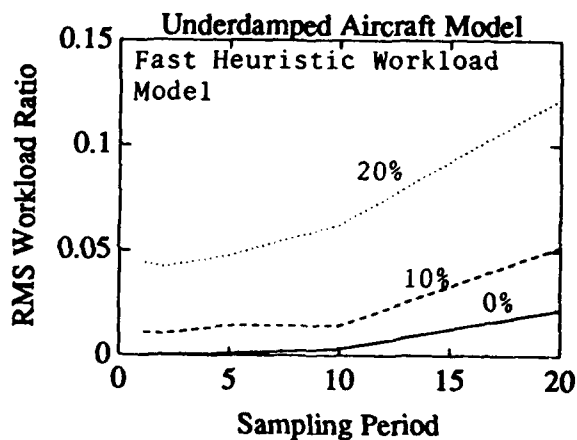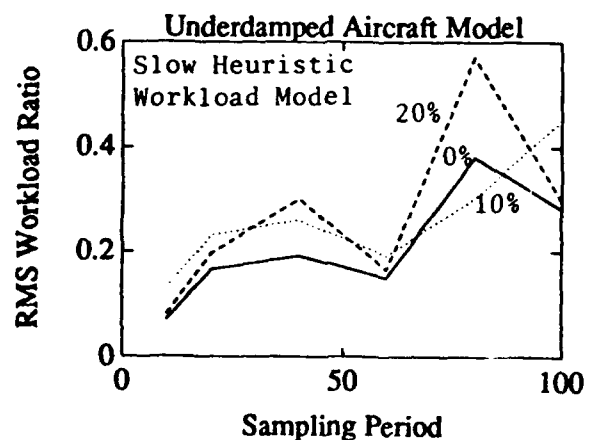Figure 9b

Figure 9c

Figure 9d

Figure 9e

Figure 9f

Figure 9. Results for Identification of Heuristic Workload Model

24

CONCLUSIONS

For linear workload models, the proposed identification model consistently performed within 10% error ratios, with noise levels up to 20% and reasonable sampling periods. Results were similar for the fast heuristic workload model, most likely because it had a strong linear component.

For the slow heuristic workload model, identification performance was predominantly in the 15-30% range. This is not surprising since a linear identification algorithm should not work as well when the data is being generated by a nonlinear process. The performance of the slow heuristic model was also less sensitive to sampling period.

Identification performance of the heuristic control model was within the 25-30% range. This performance was also fairly insensitive to sampling period, again apparently due to the lack of underlying linear dynamic relationships.

Overall, the linear identifier worked quite well for linear workload models whose inputs were a nonlinear transformation of system state and whose outputs were corrupted by noise. The discrete-continuous-discrete transformation enabled dealing with moderate sampling periods. However, at some point, the identifier was simply not getting enough information to do a good job.

The linear identifier did not work quite as well for the heuristic workload models, particularly if the linear component of its outputs was not dominant. However, its identification performance was not that bad, and it was less sensitive to sampling period. A similar conclusion can be reached for the identification of the heuristic control model.

Thus, the proposed approach to modeling the dynamics of mental workload has been shown to perform acceptably and is reasonably robust. The next step is to run similar experiments with real human controllers. If comparable results are found, the likely subsequent step will be to consider more complicated task environments.

At some point, much more sophisticated identification algorithms are likely to be needed. While such may be available and yield improved performance, they inevitably will be more complex than simple linear algorithms. In this report, we have shown that these simple algorithms perform remarkably well when the underlying "ground truth" has substantial nonlinear and heuristic components.

The practical implication of these conclusions is that online identification and prediction of workload and subsequent behaviors may be feasible. Perhaps surprisingly, very simple approaches performed adequately when the noisy, nonlinear "ground truth" differed substantially from the assumptions that typically underlie the use of these approaches. If further research, particularly human-in-the-loop empirical studies, shows that these simple concepts and their extensions work in more complex task environments, then it may be possible to substantially enhance the adaptability of aiding and decision support systems.

REFERENCES

Enstrom, K.D., & Rouse, W.B. (1977). Real time determination of how a human has allocated his attention between control and monitoring tasks. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-7*, 153-161.

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors, 26*, 519-532.

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K.R. Boff, L. Kaufman, & J.P. Thomas, (Eds.), *Handbook of perception and performance* (Chap. 41). New York: Wiley.

Hart, S.G. (1988). Workload: Where are we and where should we go next? Presented at *Workload, Training, and Performance in Advanced-Technology Systems*, Technical University of Berlin, June 10-12.

Johannsen, G., & Rouse, W.B. (1983). Studies of planning behavior of aircraft pilots in normal, abnormal, and emergency situations. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 267-278.

Lewis, C.M. (1986). *Identification of rule-based models*. Ph.D. Dissertation. Atlanta, GA: Georgia Institute of Technology.

Moray, N. (Ed.), (1979). *Mental workload: Its theory and measurement*. New York: Plenum Press.

Morris, N.M., & Rouse, W.B. (1988). *Human operator response to error-likely situations in complex engineering systems* (Tech. Report 177484). Moffett Field, CA: NASA Ames Research Center.

O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload assessment methodology. In K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and performance* (Chap. 42). New York: Wiley.

Rouse, W.B. (1977). A theory of human decision making in stochastic estimation tasks. *IEEE Transactions on Systems, Man, and Cybernetics, 7*, 274-283.

Rouse, W.B. (1980). *Systems engineering models of human-machine interaction*. New York: North Holland.

Rouse, W.B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30*, 431-443.

Rouse, W.B., Geddes, N.D., & Curry, R.E. (1988). An architecture for intelligent interfaces: Outline of an approach to supporting operators of complex systems. *Human-Computer Interaction, 3*, 87-122.

Rouse, W.B., Geddes, N.D., & Hammer, J.M. (1990). Computer-aided fighter pilots. *IEEE Spectrum, 27*, 38-41.

Rouse, W.B., Hammer, J.M., & Lewis, C.M. (1989). On capturing humans' skills and knowledge: Algorithmic approaches to model identification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-19*, 558-573.

Rouse, W.B., & Morris, N.M. (1987). Conceptual design of a human error tolerant interface for complex engineering systems. *Automatica, 23*, 231-235.